UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

Dimensionality Reduction Techniques for the Analysis of Physiological Signals in Human Sexual Response

Relatore
Prof. Varagnolo Damiano

Laureanda
Chinello Angela
Matricola: 2073948

ANNO ACCADEMICO 2024-2025

Data di laurea 23.09.2025

Summary

This thesis explores the physiological signals that underlie the sexual response of women, focusing on the identification of the orgasm phase. Using data collected at Padova SexLab, Principal Component Analysis (PCA) was applied to reduce dimensionality, highlight latent structures, and investigate differences between experimental phases.

The analysis revealed participant-specific heterogeneity, but also showed clustering patterns related to the experimental phases, suggesting that PCA can capture meaningful variations in physiological responses. Additional exploration, combining data from the same phase among participants, indicated shared patterns, emphasizing the potential of multivariate analysis to uncover consistent physiological trends.

In general, the study provides preliminary information on the sexual physiology of women and demonstrates the utility of PCA as an exploratory tool for future research.

Contents

List of Figures

Chapter 1

Introduction

1.1 Context

The study of the sexual response of women is a field of great interest, both for scientific research and for its potential clinical and therapeutic implications. Unlike other physiological functions, no objective markers have yet been identified that can distinguish the different stages that lead to orgasm in women. This shortcoming limits the development of clinical and therapeutic tools capable of deeper investigation of the physiological mechanisms underlying female sexuality. The first systematic investigations date back to the pioneering work of Masters and Johnson (1966) [1], who, analyzing a sample of men and women, described the physiological variations observed during sexual activity and proposed a linear model of sexual response. This framework was later expanded by Kaplan (1979) [2], who introduced the triphasic model (desire, arousal, orgasm). More recently, Basson (2000, 2001) [3], [4] proposed a circular model of the female sexual response, emphasizing the complex interplay between physiological, psychological, and relational factors. However, despite these conceptual advances, objective physiological indicators capable of reliably identifying the orgasmic phase remain elusive, and only a limited number of studies have addressed this topic with systematic experimental methodologies.

The project described in this paper is part of this research line. Data collection was carried out with the support of the Padova Sex Lab, an academic research group affiliated with the Department of Developmental and Social Psychology of the University of Padua, dedicated to the study of human sexuality. The laboratory was founded by Jeff Kiesner in association

with Celeste Bittoni, a doctoral student. Its activities and projects include scientific research on female sexuality, the analysis of sexuality as a subjective and social experience, scientific dissemination, and the implementation of educational and community initiatives.

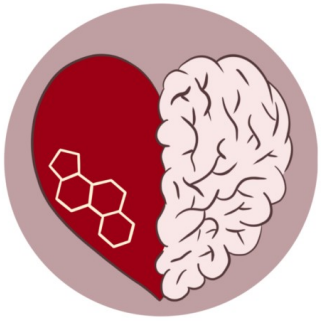


Figure 1.1: Padova SexLab's logo

The analysis presented in this paper aims to investigate whether specific physiological signals can provide significant clues to discriminate between the different phases of the female sexual response, with particular reference to the identification of the orgasm phase. My contribution is focused on the processing and interpretation of data collected, through the application of Principal Component Analysis (PCA), in order to evaluate its predictive capacity and informative potential in characterizing physiological responses.

1.2 Thesis structure

This paper is organized as follows. Chapter 2 introduces the underlying theoretical concepts, with particular attention to the physiology of the signals considered. Chapter 3 describes the data collection process for the data set and the instrumentation used. Chapter 4 discusses the analysis results and the contribution of PCA. Finally, Chapter 5 presents the conclusions.

Chapter 2

Theoretical Background

This chapter outlines the theoretical and scientific foundations necessary for analyzing the female sexual response. After a general overview of physiological signals, the selected ones used in the application of Principal Component Analysis (PCA) will be described.

2.1 Physiological Signals

Physiological signals, in both men and women, represent the expression of normal biological processes and reflect the functioning of organs and systems of the human body. They also constitute the means through which the organism perceives stimuli and interacts with the external environment. In this study, these signals were detected using biosensors that convert biological reactions into measurable data. Among the many possible physiological measures, the following were selected because of their relevance to the phenomenon under investigation:

2.1.1 Electrocardiogram

The electrocardiogram (ECG) records the electrical activity of the heart and provides information on heart rate and heart rate variability (HRV). Cardiac activity is detected by placing electrodes in specific areas of the body where the electrical signals are strongest, thus obtaining a graphical trace useful for assessing the heart's health and possible abnormalities. These parameters, influenced by the autonomic nervous system, can reflect states of arousal or relaxation. In the present study, the ECG was considered relevant because variations in HRV can reflect changes in autonomic balance associated with sexual arousal.

2.1.2 Electrodermal Activity

Electrodermal Activity (EDA), also referred to as Galvanic Skin Response (GSR), measures changes in the electrical conductance of the skin influenced by sweat gland activity. This parameter is widely used to monitor physiological arousal in response to emotional or sexual stimuli. When the sympathetic nervous system stimulates the eccrine sweat glands, the resulting increase in sweat on the skin surface improves skin conductance, which can be detected using electrodes. EDA is therefore a sensitive indicator of physiological arousal and emotional reactivity. In this study, EDA was chosen to capture variations related to sexual stimulation.

2.1.3 Respiratory Inductance Plethysmography

Respiratory Inductance Plethysmography (RIP) monitors thoracoabdominal movements during the respiratory cycle, providing information on rhythm and depth of breathing. Breathing, regulated by the nervous system, is a sensitive indicator of relaxation or arousal states. RIP records both inhalation, characterized by chest expansion and air intake, and exhalation, the phase of relaxation in which air is expelled. In the present study, RIP was used because breathing patterns are known to change in correspondence with emotional and sexual arousal.

To analyze these physiological signals jointly and extract meaningful patterns, multivariate statistical methods are required. Among these, PCA was selected for its ability to reduce the dimensionality while retaining most of the original information.

2.2 PCA: Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique widely used in multivariate statistical analysis and machine learning. Its goal is to transform a set of correlated variables into a new set of uncorrelated variables, called principal components, obtained as orthogonal linear combinations of the original variables. The principal components are ordered based on the amount of variance explained, so that the first progressively captures the largest portion of the information contained in the data. This methodology allows the data set to be projected into a reduced-dimensional space, preserving its intrinsic structure as much as

possible.

PCA is frequently used as a preprocessing step to mitigate noise and multicollinearity, in data compression, and for the visualization and identification of latent patterns or clusters. For example, if two variables are highly correlated, PCA replaces them with a single principal component, thus reducing redundancy.

Despite its advantages, PCA also has limitations: being a linear transformation, it cannot model nonlinear relationships between variables, and the interpretability of principal components can be challenging when they derive from combinations of many variables.

2.2.1 PCA Application Procedure

In this study, PCA was applied to a previously collected data table, represented as a matrix $X \in \mathbb{R}^{n \times p}$, with n observations and p variables. The main steps applied to a matrix X are the following:

1. **Data Standardization**: Standardization transforms the variables to uniform measurement scales by rescaling them to zero mean and unit variance. This prevents variables with larger ranges from dominating the analysis. Given a matrix $X = [x_{ij}]$, the standardized value is computed as

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j},$$

where μ_j and σ_j are the mean and standard deviation of column j. The resulting matrix Z has columns with mean 0 and variance 1.

2. **Computation of the Covariance Matrix**: The covariance matrix Σ summarizes the linear relationships between variables, with variances on the diagonal and covariances off-diagonal. It is defined as:

$$\Sigma = \frac{1}{n-1} Z^{\top} Z.$$

3. **Eigenvalues and Eigenvectors**: The eigenvalues λ and eigenvectors v of Σ are then computed:

$$\Sigma v = \lambda v.$$

Eigenvectors represent directions of maximum variance, and eigenvalues quantify the variance captured in each direction.

4. **Selecting Principal Components and Projecting Data**: The eigenvalues are sorted in descending order. The variance explained by each component is

$$\text{VE}_j = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i},$$

and cumulative variance is

$$\text{Cumulative VE}_k = \sum_{j=1}^{k} \text{VE}_j .$$

The first k eigenvectors form the projection matrix $W = [v_1, \ldots, v_k]$, and the standardized data are projected as:

$$S = ZW,$$

where S contains the scores of the observations in the reduced space.

2.2.2 Principal Components

Each principal component is a linear combination of the original variables and the components are ordered according to the amount of variance that they explain. The orthogonality of the components ensures that each captures unique information, avoiding redundancy and facilitating dimensionality reduction and data visualization.

- **PC1 (First Principal Component)**: captures the maximum variance in the data set. It represents the dominant pattern of variability and defines the primary axis of variation.
- **PC2 (Second Principal Component)**: orthogonal to PC1, it captures the second largest variance, providing complementary information independent from PC1.

Higher-order components (PC3, PC4, …), while explaining progressively smaller portions of variance, may still reveal subtle but meaningful patterns in the data, depending on the context of analysis. This hierarchical structure makes PCA a versatile tool for exploring complex multivariate data sets.

Chapter 3

Methods and Data Acquisition

In this chapter, the software and the experimental environment used are described. The data acquisition process is also explained, carried out on a sample of 11 healthy female participants.

3.1 Hardware and Software

This section describes the biosensors used to record physiological signals and the software employed to manage data collection and subsequent analysis.

3.1.1 BiosignalsPlux

BiosignalsPlux is a multisensor platform produced by PLUX Biosignals (a Portuguese technology company founded in 2007), designed for the acquisition of physiological signals in research and development contexts [5].

In the present study, the following signals were measured using this biosensor: Electrocardiogram (ECG), Electrodermal Activity (EDA), and Respiratory Inductance Plethysmography (RIP).

Figure 3.1: BiosignalsPlux device used for multimodal physiological signal acquisition.

This device has eight input channels (each of which can measure a different physiological signal). In our case, the following sensor cables were connected:

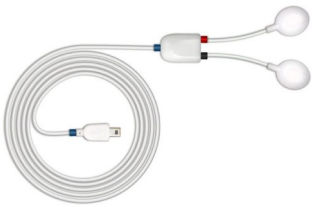


Figure 3.2: EDA cable used for electrodermal activity recording.

Figure 3.3: RIP cable used for respiratory inductance plethysmography recording.

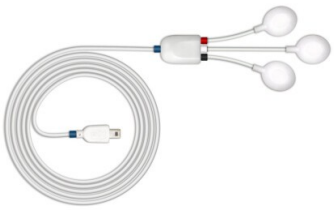


Figure 3.4: ECG cable used for electrocardiogram recording.

3.1.2 iMotions

iMotions is a software platform designed for data collection and analysis that integrates multiple physiological and behavioral sensors. In the present study, it was used to synchronize and store the data recorded by biosensors [6].

Figure 3.5: Logo of the iMotions software platform.

3.1.3 MATLAB

MATLAB, developed by MathWorks, is a software environment and programming language widely used for numerical computation, data analysis, algorithm development, and visualization. [7].

Figure 3.6: Logo of MATLAB (MathWorks).

3.2 Data Acquisition Phase

The data acquisition phase was structured according to precise methodological steps to ensure consistency and comparability among the participants. For each subject, a baseline was first identified under controlled conditions; it served as reference points for the normalization of physiological signals, to reduce interindividual variability and enable meaningful intra- and intersubject comparisons:

- Breathing exercises for the RIP signal;

- Kegel exercises for CH0 and CH1 channels (in this study, intravaginal pressure was recorded as a physiological signal. However, due to the complexity of the data and difficulties in interpretation, this variable was not included in the analyses presented here and is therefore only mentioned in this section);
- Resting state for ECG and EDA.

During the experimental task (exposure to pornographic material), the recordings were segmented into four distinct phases: ***pre-stimulation***, ***stimulation***, ***peak*** (orgasm), and ***post-stimulation***. For each phase, the analysis was performed over time windows of equal length, defined by the duration of the peak event. This segmentation allowed a homogeneous comparison between the different stages, with particular emphasis on the peak interval, which was considered the most relevant for the identification of physiological responses.
Both intra-subject and intersubject analyses were performed. In the first case, the features extracted from the signals of each participant were normalized with respect to their baseline, enabling the characterization of individual responses. In the second case, the events between participants were compared by averaging the time unit t associated with the peak and verifying the recurrence of the physiological event at the expected intervals.

3.2.1 Technical Considerations

The choice of time unit t was critical: while the peak lasted approximately 10 seconds, this interval could be insufficient for signals with higher variability, such as RIP and EDA. For this reason, adjacent intervals preceding and following the peak were also considered. Particular attention was paid to possible **overflow phenomena**, characterized by abrupt jumps in signal amplitude, which required recalibration and cross-validation with manufacturer data. The quality of the **signal** was another key issue, as noise and potential missing peaks were observed, especially in the EDA signals, probably influenced by the low ambient temperature during data collection.
This methodological framework ensured that the acquired data could be reliably normalized, segmented, and compared, providing a solid foundation for the subsequent extraction and statistical analysis of the features.

Chapter 4

PCA Analysis

In this chapter, we present the analysis conducted on the extracted data, following the methodology introduced in the previous section. Principal Component Analysis (PCA) was employed as a dimensionality reduction technique and as an exploratory tool for the physiological signals under investigation.

The implementation was carried out in the **MATLAB** environment, through dedicated scripts that generated several visualizations, each aimed at highlighting different aspects of the data:

- The **scores plot** projects the observations onto the principal components, with each point representing a sample (e.g., subject or experimental phase) in the reduced space. This visualization is particularly useful for identifying clusters or separations that may reflect meaningful physiological distinctions.

Figure 4.1: Example of a scores plot

- The **loadings plot** illustrates the contribution of each original variable to the principal components. It provides insight into which physiological signals account for the variance captured by each component, thus supporting the interpretation of the underlying dimensions.

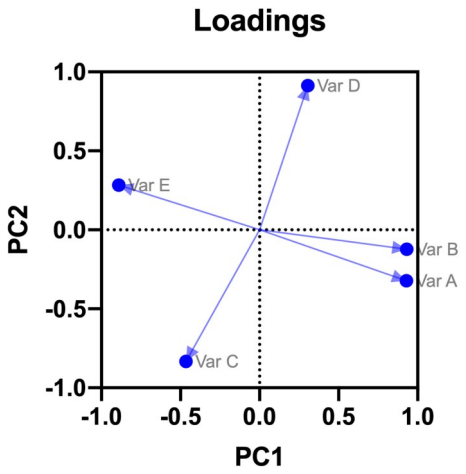


Figure 4.2: Example of a loadings plot

- The **explained variance plot** displays the proportion of total variance in the dataset explained by each principal component. This information is crucial for determining the number of dimensions sufficient to represent the data without excessive information loss.

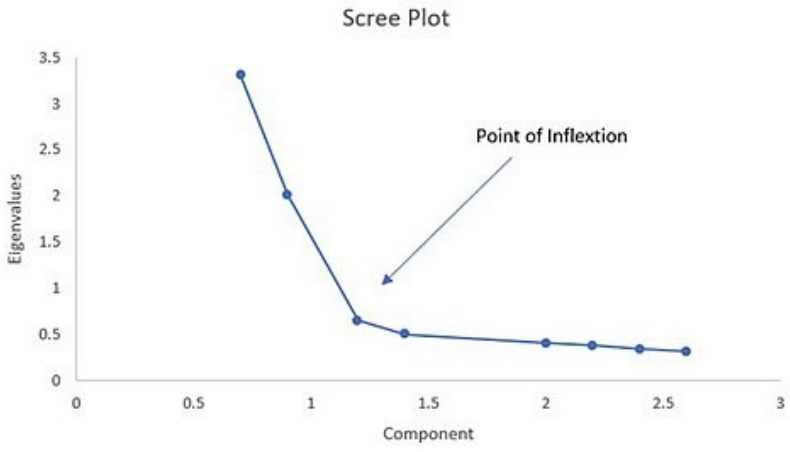


Figure 4.3: Example of an explained variance (scree) plot

Among these, particular emphasis was placed on the scores plot, as it was hypothesized to contain the most relevant information for the purposes of this study. Specifically, the score

representation was explored as a potential tool for identifying, and possibly predicting, the phase corresponding to orgasm.

For each participant, a dedicated folder was created to store the collected data. As noted previously, the data was cleaned as thoroughly as possible; however, they remained heterogeneous between participants. In this context, Principal Component Analysis (PCA) was selected because of its ability to reduce dimensionality and highlight latent structures within complex multivariate datasets such as physiological signals. By projecting the original variables onto a smaller number of components, PCA facilitates the identification of underlying patterns that may not be immediately visible, thus allowing the exploration of potential relationships between the different phases of the sexual response.

The analysis also had to account for specific cases, such as one participant who did not reach the orgasm phase and another who experienced it twice. These circumstances, combined with the limited number of participants, inevitably reduced the precision and statistical significance of the results. However, such individual variability is precisely the type of heterogeneity that PCA is designed to manage, as it captures shared components while attenuating the influence of noise or subject-specific anomalies. In a larger sample, these individual differences would probably have had only a marginal impact on the overall findings.

4.1 Code Example

The following MATLAB script illustrates the procedure used to perform PCA on the collected data. Its purpose is to standardize the physiological signals, compute the principal components, and generate the corresponding visualizations for each participant. Each table of data contains columns representing the different physiological signals and rows corresponding to the experimental phases described in Chapter 3.

4.1.1 Loading the Data

The first step consists of importing the participant's dataset from a CSV file into MATLAB.

```
dataTable = readtable('NAME. OF . THE . FILE / FILE . PATH');
```

This operation allows for subsequent processing of the data matrix in a consistent format for all participants.

4.1.2 Matrix Preparation and Handling Missing Values

Next, numerical columns are selected, and missing values (NaN) are replaced with zeros to avoid computational errors during PCA.

```
numericCols = varfun(@isnumeric, dataTable, 'OutputFormat', 'uniform');
dataMatrix = table2array(dataTable(:, numericCols));
dataMatrix(isnan(dataMatrix)) = 0;
```

4.1.3 Data Standardization

Standardization is performed to center the data and scale the variables to unit variance, as discussed in Chapter 2. This ensures that each variable contributes equally to the principal components.

```
mu = mean(dataMatrix, 2);
sigma = std(dataMatrix, 0, 2);
dataStandardized = (dataMatrix - mu) ./ sigma;
```

4.1.4 PCA Computation

Principal Component Analysis is performed using singular value decomposition (SVD) to compute the eigenvectors, scores, and explained variance.

```
mu = mean(dataMatrix, 1);
dataCentered = dataMatrix - mu;
[U, S, V] = svd(dataCentered,'econ');
coeff = V;          % Principal coefficients (eigenvectors)
score = U * S;      % Projected data onto principal components
latent = diag(S).^2 / (size(dataMatrix,1) - 1); % Eigenvalues
explained = 100 * latent / sum(latent);
% Explained variance
```

4.1.5 Visualization of Results

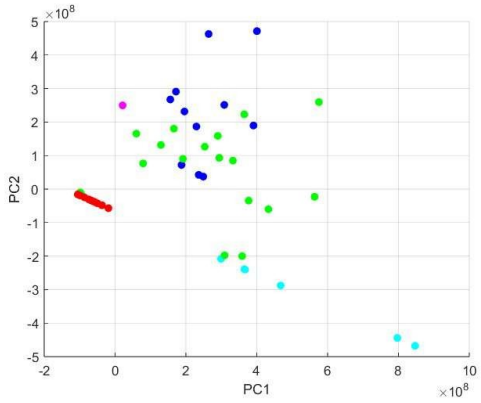
After computing the principal components, visualizations, such as score plots, loadings plots, and explained variance (scree) plots, are generated for each participant. These plots allow inspection of potential clusters, the contribution of each physiological signal to the principal components, and the proportion of variance captured by each component.

4.2 Results

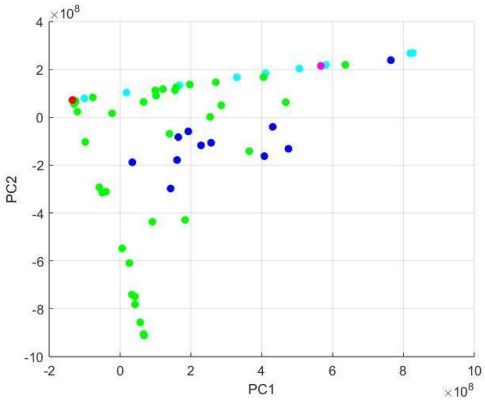
For each of the 11 participants, a scores plot was generated, representing the projection of the experimental phases onto the first two principal components (PC1 and PC2). Each point in the plots corresponds to a time window, color-coded according to the experimental condition described as follows:

- Resting cross cut
- Kegel exercises
- Eyes closed
- Porn phase
- Breathing cut
- Breathing cut post

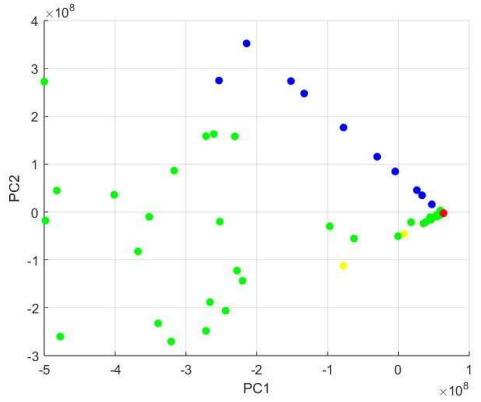
The following images report the complete set of visualizations obtained.



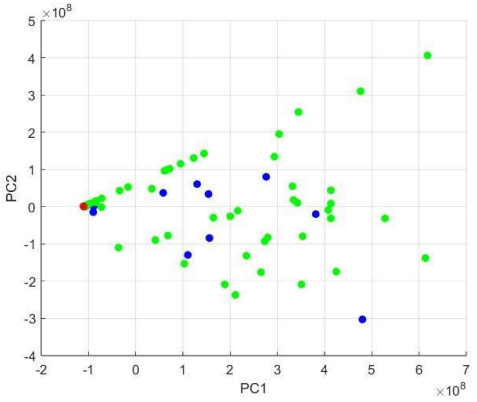
Sample 1



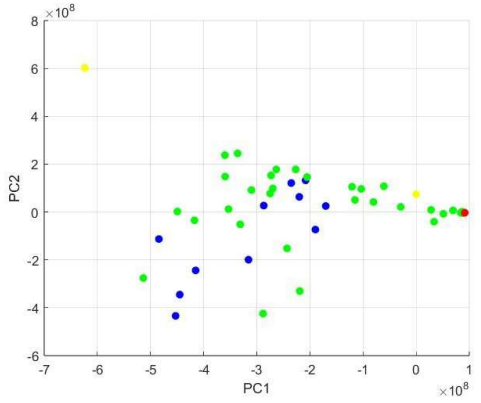
Sample 2



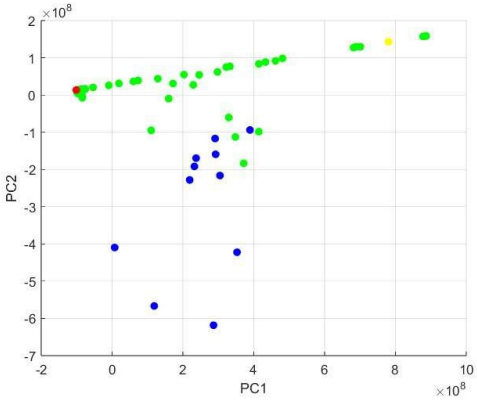
Sample 3



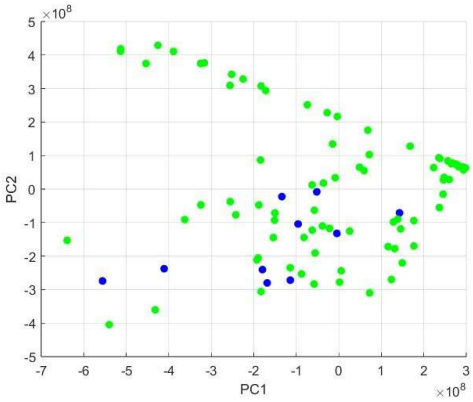
Sample 4



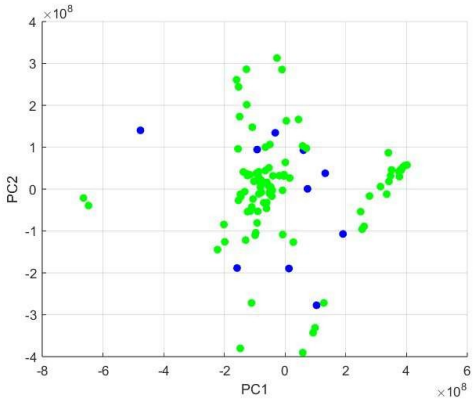
Sample 5



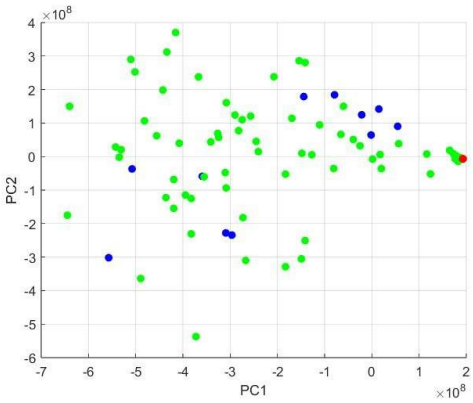
Sample 6



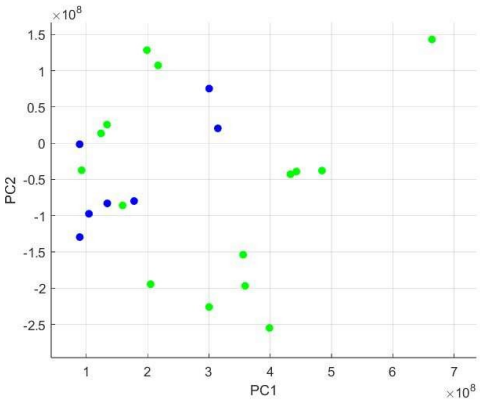
Sample 7



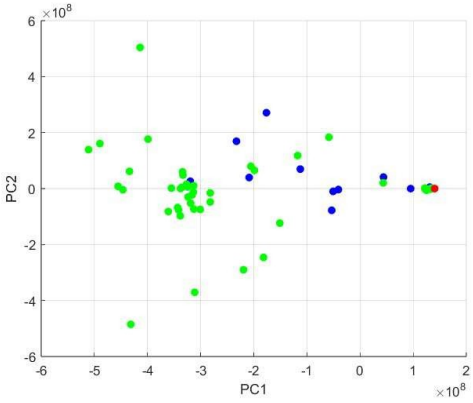
Sample 8



Sample 9



Sample 10



Sample 11

From the visual inspection of the plots, some general patterns can be observed:

- **Cluster formation**: in several participants, points corresponding to the same experimental phase tend to cluster together, suggesting that PCA captures differences in physiological signals between phases.

- **Phase separation**: in particular cases, the pornographic stimulation phase shows a displacement along one of the principal components compared to the resting or breathing phases. This suggests that the physiological response to sexual arousal may be represented in a reduced-dimensional space.
- **Heterogeneity across participants**: the distribution and orientation of the groups vary substantially between participants. This variability is likely due to individual differences, signal noise, and the small sample size.

In general, while no perfectly consistent pattern emerged in all participants, the plots indicate that PCA can highlight phase-related differences in physiological signals. In some cases, the transition towards the orgasm phase appears as a trajectory in the scores plot, with a higher concentration of points around the peak followed by a dispersion during the post-stimulation period.

4.3 Remarks

The visualizations obtained for the 11 participants highlight the heterogeneity of the results, both in terms of point distribution and signal patterns. These observations emphasize the exploratory nature of this analysis and the need for cautious interpretation.

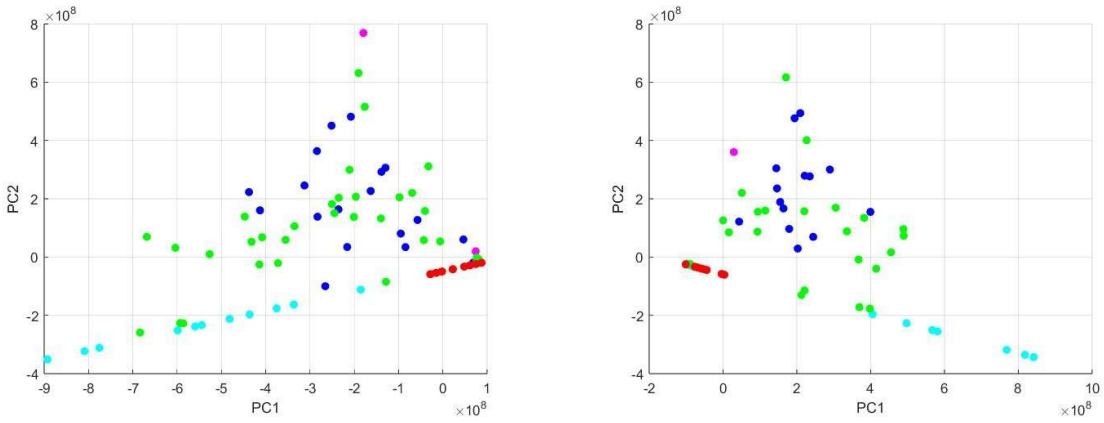
The score plots vary between participants, not only in the point distribution but also in the orientation of the first two principal components (PC1 and PC2). Initially, time intervals of 10 seconds were considered; shorter intervals of 5 and 8 seconds were tested to evaluate their effect on the results.

For Participant 1, shorter intervals result in a more fragmented and heterogeneous representation, as reflected by the increased dispersion of colored points. This illustrates the higher variability of physiological signals in shorter time frames.

Preliminary observations show that certain regions of the plots exhibit higher point densities, which gradually decrease moving away from what is hypothesized as the peak corresponding

to the orgasm phase. Although this trend is not consistent across all participants, it suggests the presence of underlying patterns that could be further explored with larger samples and complementary analytical methods.

The following comparison for Participant 1 illustrates the effect of different time intervals:



5 seconds 8 seconds

It is evident that smaller intervals produce significant changes in the plots, highlighting greater variability and finer temporal details in the physiological signals.

An additional analysis was performed by selecting, for all participants, the first two columns of the original data table corresponding to the porn phase. These columns were combined among participants and PCA was applied to this aggregated dataset to generate a new scores plot.

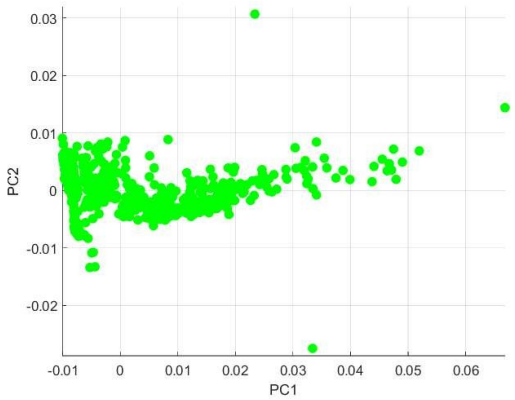


Figure 4.4: Scores plot resulting from the combined analysis.

The resulting plot shows that the points cluster in a single region, forming a compact distribution with a coherent trend. This indicates that aggregating data from the same phase between participants reveals shared patterns, highlighting the overall shape of the physiological response during pornographic stimulation.

Chapter 5

Conclusion

The objective of this analysis, as mentioned at the beginning, was to establish a systematic method to manage, process, and compare data, highlighting critical issues and the need for further technical investigations. The idea was to use data collection and analysis to, in some way, predict the peak determined by the achievement of orgasm.

Since this project is still in its early stages and deals with a new area, without historical data available from previous decades, it has not yet been possible to reach definite and reliable conclusions. However, the work carried out has confirmed the relevance of developing quantitative methods to analyze physiological data related to sexual response, a field that is still little explored but with significant potential both in scientific research and in practical applications.

In practical terms, my contribution allowed us to test whether this data analysis technique, specifically PCA, could provide any meaningful response. More importantly, it laid the groundwork for the application of structured data analysis approaches to this domain, providing a first framework on which future research can build. Due to limitations such as the small number of participants and incomplete or imperfectly recorded data (including the two special cases among participants), the results obtained did not yield particularly relevant information.

5.1 Future Work

Given the early stage of the project, more time will be needed to achieve more robust answers. Possible directions for future work include the following.

- Increasing the number of collected samples to allow for analysis on a larger scale and obtain more robust results.
- Exploring alternative data analysis techniques different from PCA to determine whether the limitations observed are due to the method itself or the quality of the data collected.
- Continuing data collection with more accuracy, reducing errors or gaps in physiological signal recording.

Although the present results are not yet conclusive, the work carried out demonstrates that the approach is promising and worth pursuing. With further development, it may ultimately contribute to a deeper understanding of human physiological and psychological processes, providing valuable insights into a phenomenon that is both complex and scientifically relevant.

Bibliography

[1] W. H. Masters and V. E. Johnson, *Human Sexual Response*. Boston: Little, Brown and Company, 1966.

[2] H. S. Kaplan, *Disorders of Sexual Desire*. New York: Brunner/Mazel, 1979.

[3] R. Basson, "The female sexual response: A different model," *Journal of Sex & Marital Therapy*, vol. 26, no. 1, pp. 51–65, 2000. doi: 10.1080/009262300278641

[4] R. Basson, "Human sex-response cycles," *Journal of Sex & Marital Therapy*, vol. 27, no. 1, pp. 33–43, 2001. doi: 10.1080/71384714 6

[5] PLUX – Wireless Biosignals S.A., *Plux biosignals*, https://www.pluxbiosignals.com/.

[6] iMotions A/S, *Imotions - human behavior research software*, https://imotions.com/

[7] MathWorks, *Matlab documentation*, The MathWorks, Inc., Natick, Massachusetts. [Online]. Available: https://www.mathworks.com/help/matlab/